A Sketching Framework for Reduced Data Transfer in Photon Counting Lidar

A Sketching Framework for Reduced Data Transfer in Photon Counting Lidar

Michael P. Sheehan, *Member, IEEE*, Julián Tachella, *Member, IEEE* and Mike E. Davies, *Fellow, IEEE*

***Abstract*—Single-photon lidar has become a prominent tool for depth imaging in recent years. At the core of the technique, the depth of a target is measured by constructing a histogram of time delays between emitted light pulses and detected photon arrivals. A major data processing bottleneck arises on the device when either the number of photons per pixel is large or the resolution of the time stamp is fine, as both the space requirement and the complexity of the image reconstruction algorithms scale with these parameters. We solve this limiting bottleneck of existing lidar techniques by sampling the characteristic function of the time of flight (ToF) model to build a compressive statistic, a so-called sketch of the time delay distribution, which is sufficient to infer the spatial distance and intensity of the object. The size of the sketch scales with the degrees of freedom of the ToF model (number of objects) and not, fundamentally, with the number of photons or the time stamp resolution. Moreover, the sketch is highly amenable for on-chip online processing. We show theoretically that the loss of information for compression is controlled and the mean squared error of the inference quickly converges towards the optimal Cramér-Rao bound (i.e. no loss of information) for modest sketch sizes. The proposed compressed single-photon lidar framework is tested and evaluated on real life datasets of complex scenes where it is shown that a compression rate of up-to 1/150 is achievable in practice without sacrificing the overall resolution of the reconstructed image.**

***Index Terms*—Single-Photon Lidar, Empirical Characteristic Function, Compressive Learning, Summary Statistics**

I. Introduction

SINGLE photon light detection and ranging (lidar) has emerged as an important depth imaging technique prevalent in the automobile [1], [2], defence [3] and forestry industries [4]. This modality has the unique advantage of offering very high depth resolution [5] even at long-range scenes using low-power (eye-safe) lasers [6]. The technique has at its core the ability of emitting light pulses and detecting each single-photon as it arrives, thereby obtaining a depth estimate by measuring the round-trip time of individual photons. By using a time correlated single-photon counting (TCSPC) system, a histogram can be formed indicating the time delay between emitted light pulses and detected photons for each pixel, with a proportion of the photons originating from background or ambient light (e.g. the sun). The number of counts per time histogram bin provide information on the depth and reflectivity of the object or scene. The presence of a peak in the histogram indicates an object is present within the range of the lidar system. The location of this object corresponds to the location of the impulse response. If the material is semi-transparent (e.g. glass, water, camouflage) or the laser footprint is large, then multiple peaks with different intensities may exist within a single pixel [5]. A standard example of a TCSPC histogram for a given pixel within a scene is shown in Figure 1.

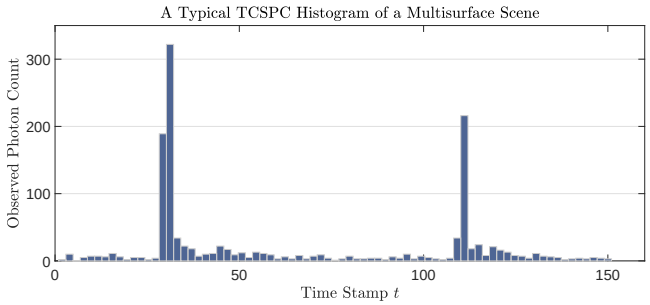

Fig. 1: An example of a TCSPC histogram of a pixel in a complex scene including a semi-transparent material (camouflage) in front of a person which is depicted by the 2 spikes of the histogram respectively.

The image restoration task reduces down to inferring the positions and intensities of the peaks in the histogram for each pixel in the image. Typically, either the time stamp of all photons or a histogram of bin width Δt has to be recorded, stored in memory and transferred from the chip for each pixel in the scene. The development of high rate, high resolution, low power ToF image sensors is challenging due to the large data volumes required. This causes a major data processing bottleneck on the device when either the number of photons per pixel n is large, the time resolution, Δt, is fine or the spatial resolution is high, as the space requirement, power consumption and computational burden of the depth reconstruction algorithms scale with these parameters [7].

Various existing methods have attempted to tackle the trade-off between depth resolution and computational/space complexities. A number of papers [8]–[12] propose methods to address the trade-off between depth resolution and the complexities associated with the TCSPC histogram. Henderson et al. [8] propose a method that employs a gated procedure to coarsely bin the detected photons, whilst Ren et al. [9] develop a sliding window approach to achieve high resolution depth. Walker et al. [10] calculate the depth directly from the photon time stamps. However in all of these approaches, the approximations formed on-chip compromise the depth resolution of the image. Della Rocca et al. [11], [12] proposes to only collect the histograms of photon detections when there is a significant change of activity. This method reduces the data-transfer, as it is only required during specific moments in time. Similarly, Hutchings et al. [13] propose a method of discarding photon

M.Sheehan, J.Tachella and M.Davies are with the School of Engineering, University of Edinburgh. This work is supported by ERC Advanced grant 694888, C-SENSE and a Royal Society Wolfson Research Merit Award. Correspondence to M.Sheehan (Email: michael.sheehan@ed.ac.uk).

detections based on activity. However, these methods can potentially remain idle when there is a small change in activity, and can also suffer from a loss of temporal resolution due to coarse histogram binning. Zhang et al. [14] propose a method of reducing the transfer of photon detections by performing a coarse to fine approximation of the histogram of time-of-flight. At each scale, a coarse histogram is constructed with a limited number of bins. Multiple histograms of increasing resolution have to be formed, hence the method has an increased total acquisition time and can also suffer from a loss of temporal resolution.

Compressive sensing strategies have been successfully applied to lidar [15]–[17], focusing on compressing the information across pixels. Kadambi et al. [16] propose to exploit the sparsity of natural scenes in some representation domain (e.g. wavelet transform) to greatly reduce signal acquisition. The depth accuracy is limited by the level of amplitude noise and decay of the impulse response and is therefore limited to the case of one surface per pixel. In a similar vein, Halimi et al. [18] propose an adaptive sampling strategy that is scene dependent. By building up regions of interest and data driven depth maps in an iterative manner, they efficiently choose suitable scan positions to reduce acquisition time by up to 8 times in comparison to uniform sampling in certain scenarios. However, these methods perform compression within the spatial domain and not, in the case of our method, throughout the depth or time domain and are therefore fundamentally different in practice to the method proposed in this paper.

Another approach to reduce the data transfer of the information needed to reconstruct the lidar image is to compress the data on-chip. As highlighted in [19], standard low-level data compression methods can be used to compress the data on-chip, however these methods can only offer up to a modest 50% data reduction and in some cases involve significant on-chip computation or there are limitations with respect to on-chip storage.

In this paper, we propose a novel solution to this bottleneck of existing lidar techniques by calculating on-the-fly summary statistics of the photon time stamps, a so-called sketch, based on samples of the characteristic function of the ToF model. Distinct to compressive sensing, the goal here is not to recover the photon counting data but rather the underlying probability distribution. In this sense, we are estimating the probability model directly from some summary statistics and therefore our proposed framework utilises much of the theory found in the generalised method of moments [20], [21], empirical characteristic function [22], [23] and compressive learning [24]–[26] literature. The size of the sketch scales with the degrees of freedom of the ToF model (i.e., number of objects in depth) and not with the number of photons or the fineness of the time resolution, without sacrificing precision in depth. The sketch can be computed for each incoming photon in an online fashion, only requiring a minimal amount of additional computation which can be performed efficiently on-chip. The sketch can be shown to capture all the salient information of the histogram, including the ability to explicitly remove background light or dark count effects, in a compact and data-efficient form, suitable for both on-chip processing or off-chip post processing. Furthermore, we develop a compressive lidar image reconstruction algorithm which has computational complexity dependent only on the size of the sketch. Our proposed method paves the way for high accuracy 3D imaging at fast frame rates with low power consumption. In summary the main contributions of the paper are as follows:

- We propose a principled approach for compressing time-of-flight information in an online fashion without the requirement to form a histogram and without compromising depth resolution.
- A compressive single-photon lidar algorithm is proposed which does not scale with either the number of photons or the time stamp resolution in terms of space and time complexity.
- The statistical efficiency, given a compression rate (or sketch size), is quantified for different single-photon lidar scenarios, showing that only limited measurements of the characteristic function are needed to achieve negligible information loss.

The remainder of the work is organized as follows. Section 2 details the ToF observation model used in single-photon lidar and also presents the idea of summary statistics used for parameter estimation including the empirical characteristic function and compressive learning statistics. In Section 3 we detail our proposed compressive single-photon lidar scheme including a reconstruction algorithm that has computational complexity which scales with the sketch size m as well as quantifying the statistical efficiency of the estimated parameters θ. Results of the compressive lidar framework are analysed on both synthetic and real datasets in Section 4. Section 5 finally summarizes our conclusions and discusses future work.

II. Background

A. Lidar Observation Model

The photon count at time stamp $t \in [0, T-1]$, for an arbitrary pixel can be modelled as a Poisson distribution [27], [28]:

$$y_{t_k}|(r, b, t_k) \sim \mathcal{P}(rh(t - t_k) + b), \tag{1}$$

where $r \geq 0$ denotes the reflectivity of the detected surface, $h(\cdot)$ is the impulse response of the system and b defines the level of background photons. The number of discretized time stamp bins over the range of interest is denoted by T. The time stamp t is discretized over the range $[0, T-1]$ dependent on the time-stamp resolution Δt. For simplicity, here we assume that the integral of the impulse response $H = \sum_{t=1}^{T} h(t)$ is constant although the proposed approach can accommodate more complex scenarios. If the lidar system is in free running mode where multiple acquisitions of a surface/object are obtained, then the interval $[0, T-1]$ can be thought of as circular in the sense that time-stamp T is equivalent to the time-stamp 0.

Alternatively, one can instead model the time of arrival of the pth photon detected. We assume there are K distinct reflecting surfaces, where α_k and α_0 denote the probability that the detected photon originated from the kth surface and background sources, respectively. Let $x_p \in [0, T-1]$ denote

the time stamp of the pth photon where $1 \leq p \leq n$, then x_p can be described by a mixture distribution [29]

$$\pi(x_p|\alpha_0, ..., \alpha_K, t_1, ..., t_K) = \sum_{k=1}^{K} \alpha_k \pi_s(x_p|t_k) + \alpha_0 \pi_b(x_p), \tag{2}$$

where $\sum_{k=0}^{K} \alpha_k = 1$. The distribution of the photons originating from the signal and background are defined by $\pi_s(x_p|t) = h(x_p - t)/H$ and the uniform distribution $\pi_b(x_p) = 1/T$ over $[0, T-1]$, respectively. Often in practice, the signal distribution π_s is modelled either using a discretized Gaussian distribution over the interval $[0, T-1]$ or through the data driven impulse function which is calculated through experiments. In Section IV, we consider both.

B. Summary Statistics

Our acquisition goal is to obtain parameter estimates of the signal model in (2), given the time-stamp of photons detected. Parameter estimation usually involves the inference of a set of parameters $\theta \in \Theta \subset \mathbb{R}^{2K+1}$ associated to a probability model $\pi(\cdot \mid \theta)$ defined on some space $\mathbf{x} \in \mathbb{R}^d$. In the case of single-photon lidar, the dimension $d = 1$. Typically, we observe a finite dataset $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^n$ of n samples which we assume is sampled from the distribution given in (2). Maximum likelihood estimation (MLE) is a traditional parameter estimation method whereby a likelihood function associated with the finite data is maximized with respect to the model parameters, e.g.

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \log \pi(\mathbf{x}_i \mid \theta). \tag{3}$$

1) Generalised Method of Moments: In some cases, the likelihood function might not have a closed form solution nor a computationally tractable approximation [20]. Generalised method of moments [20], [21] (GeMM) is an alternative parameter estimation method where one estimates θ by matching a collection of generalised moments with an empirical counterpart computed over a set of finite data sampled from the distribution $\pi(\mathbf{x} \mid \theta)$. Given a nonlinear function $g : \mathbb{R}^d \to \mathbb{C}^m$, then we define the expectation constraint

$$\mathbb{E}g(\mathbf{x}; \theta) = 0, \tag{4}$$

where $\mathbb{E}$ denotes the expectation with respect to the probability distribution $\pi(\mathbf{x} \mid \theta)$. Typically, the GeMM estimator is obtained by minimising a quadratic cost of the empirical discrepancy with respect to θ to try impose the moment constraints of (4). Let us define

$$g_n(\mathcal{X}; \theta) \coloneqq \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{x}_i; \theta), \tag{5}$$

calculated over $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^n$, then a GeMM classically takes the form [20], [21]

$$\hat{\theta} \coloneqq \arg\min_{\theta} g_n(\mathcal{X}; \theta)^T \mathbf{W} g_n(\mathcal{X}; \theta), \tag{6}$$

where $\mathbf{W}$ is a symmetric positive definite weighting matrix that may depend on θ.

2) Compressive Learning: Building on the concept of GeMM, compressive learning [24], [25] utilises generalised moments of the data but with the distinct goal of reducing signal acquisition, space and time complexities. The link to GeMM is established by separating the function g into the following particular form:

$$g(\mathbf{x}; \theta) = \Phi(\mathbf{x}) - \mathbb{E}_\theta \Phi(\mathbf{x}), \tag{7}$$

where $\Phi : \mathbb{R}^d \mapsto \mathbb{C}^m$ is often referred to as the feature function. The separable form decouples the measured moments, $\Phi(\mathbf{x})$, from the parameters θ that are to be estimated. This is not a usual assumption in GeMM, although it may arise in particular cases. By denoting the empirical mean or the so-called sketch as

$$\mathbf{z}_n \coloneqq \frac{1}{n} \sum_{i=1}^{n} \Phi(\mathbf{x}_i), \tag{8}$$

we can estimate θ solely from the sketch $\mathbf{z}_n$ by minimising

$$\hat{\theta} = \arg\min_{\theta} \|\mathbf{z}_n - \mathbb{E}_\theta \Phi(\mathbf{x})\|_{\mathbf{W}}^2, \tag{9}$$

which is the particular compressive GeMM loss of (6). In Section III, we explicitly define the weighting matrix $\mathbf{W}$ for compressive single-photon lidar.

The separable form of g in (7) allows a sketch statistic $\mathbf{z}_n$ to be formed with a single pass of the data without the need to store $\mathcal{X}$, and it can easily be updated on the fly with minimal computational cost. The sketch statistic has size m, or size $2m$ if decoupled into its real and imaginary components, which, fundamentally, scales independent of the dimensions of the dataset $\mathcal{X}$, which in the case of single-photon lidar is the photon count n or the binning resolution T.

3) Empirical Characteristic Function: A specific type of GeMM is the empirical characteristic function (ECF) estimation [22], [23], [30], and occurs when the generalized moment is chosen to be $\Phi(\mathbf{x}) = [e^{\mathrm{i}\omega_j^T \mathbf{x}}]_{j=1}^m$, where $\mathrm{i} = \sqrt{-1}$ and ω_j is a discrete set of frequencies. It is of particular interest as the expectation of Φ, namely $\Psi_\pi(\omega) = \mathbb{E}_\theta e^{\mathrm{i}\omega^T \mathbf{x}}$, is specifically the characteristic function (CF) of the probability distribution $\pi(\mathbf{x} \mid \theta)$ at frequency ω. The CF exists for all distributions, and often has a closed form expression. Moreover, it captures all the information of the probability distribution [31], therefore giving a one-to-one correspondence between the CF and the probability distribution $\pi(\mathbf{x} \mid \theta)$. The CF also has the favourable property that it decays in frequency, i.e. $\Psi_\pi(\omega) \to 0$ as $\omega \to \infty$, under mild conditions on the probability distribution $\pi(\mathbf{x} \mid \theta)$ [31], [32].

For a single depth observation model in (2) (i.e. $K = 1$) and a discrete impulse response function h, we define the characteristic function of the observation model by

$$\begin{aligned} \Psi_\pi(\omega) =& \alpha_1 \Psi_{\pi_s}(\omega) + \alpha_0 \Psi_{\pi_b}(\omega) \\ =& \alpha_1 \hat{h}(\omega) e^{\mathrm{i}\omega t} + \alpha_0 \operatorname{sinc}(\omega T/2) \end{aligned} \tag{10}$$

where $\operatorname{sinc}(x) = \frac{\sin(x)}{x}$ and $\hat{h}$ denotes the (discrete) Fourier transform of the impulse response function h. It should be noted that we could consider different distributions π_b, and hence CFs, to model the detected photons originating from

more complex background sources, although this is beyond the scope of this paper.

The feature function Φ is a complex valued function of size m. With regards to hardware implementation, it is often preferable and convenient to work directly with real valued functions. The complex term $e^{\mathrm{i}\omega x}$ can be alternatively written as $\cos(\omega x) + \mathrm{i}\sin(\omega x)$, where $e^{\mathrm{i}\omega x}$ has been decoupled into its real and imaginary components. As a result, the feature function Φ can be equivalently written as a real valued feature function $\Phi : \mathbb{R}^d \mapsto \mathbb{R}^{2m}$, consisting of $2m$ real valued terms by stacking the real and complex components, for e.g.

$$\Phi(x) = \begin{bmatrix} \cos(\omega_1 x) \\ \vdots \\ \cos(\omega_m x) \\ \sin(\omega_1 x) \\ \vdots \\ \sin(\omega_m x) \end{bmatrix}.$$

For sake of fair comparison to existing hardware implemented methods in the literature, the results and figures presented represent a sketch of size $2m$, consisting of $2m$ real valued measurements. The nature of the feature function, in terms of it being represented by a complex or real valued function, will be made clear in its context throughout the paper.

III. Sketched Lidar

We start with a warm up example to highlight the potential of using a sketch for single-photon lidar and to motivate the design of the sketch sampling procedure which will be discussed in Section III-B1.

A. Compressing Single Depth Data

In the absence of photons originating from background sources and the presence of a single surface or object, the sample mean of all the photon time-stamps ($\Phi(x) = x$) is the simplest summary statistic for estimating the single location parameter t_1. This only holds in the noiseless case as the sample mean estimate is heavily biased toward the centre of the histogram when background photons are detected.

Suppose, we instead observe the cosine and sine of each photon count x with angular frequency $\omega = \frac{2\pi}{T}$, namely

$$\Phi(x) = \begin{bmatrix} \cos\left(\frac{2\pi x}{T}\right) \\ \sin\left(\frac{2\pi x}{T}\right) \end{bmatrix}, \tag{11}$$

and denote $\mathbf{z}_n$ the real valued sketch of size 2 ($m = 1$) computed over the dataset $\mathcal{X}$ as in (8). It is possible to recover an estimate of the single depth location parameter t_1 directly from the sketch, without recourse to the data $\mathcal{X}$, via the trigonometric sample mean

$$\hat{t}_1 = \frac{T}{2\pi} \arg\left\{ \sum_{j=1}^{n} \cos\left(\frac{2\pi x_j}{T}\right) + \mathrm{i} \sum_{j=1}^{n} \sin\left(\frac{2\pi x_j}{T}\right) \right\} \tag{12}$$

where arg is the complex argument. As the background photons are distributed uniformly over the interval $[0, T-1]$ ($\pi_b(x) = \frac{1}{T}$), the expected moment of the photons originating from background sources is zero, $\mathbb{E}_{x\sim\pi_b}\Phi(x) = \mathbf{0}$. The resulting estimate of the single depth parameter $\hat{t}_1$ is therefore an unbiased estimator of the location parameter t_1. The estimator in (12) coincides with the circular mean estimator detailed in [33].

We summarise the above using a simulated example, where a pixel of $T = 1000$ histogram bins with a signal-to-background ratio (SBR) of 1 and a total of $n = 600$ photons is simulated, where the time stamp of each photon is denoted by $\mathcal{X} = \{x_i\}_{i=1}^n$. The data was simulated using a Gaussian impulse response function with $\sigma = 15$ and a true position at time stamp $t_1 = 320$. Computing the sketch $\mathbf{z}_n$ from (8) and using (12) we obtain the sketch estimate $\hat{t}_{\text{cm}} = 323.3$ and the sample mean estimate of $\hat{t} = 434.1$. The TCSPC histogram along with both the circular and standard mean estimates as well as the location parameter t_1 are shown in Figure 2 where it is evident that the circular mean estimate does not suffer from the noise bias inherent in the sample mean.

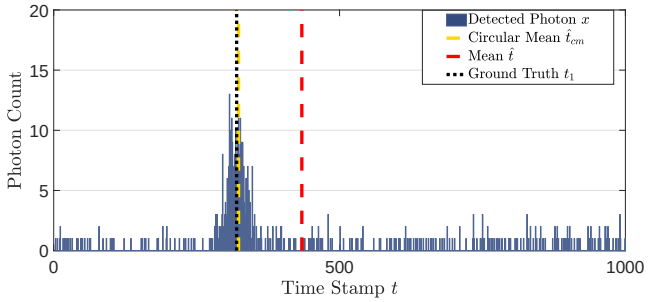

Fig. 2: The TCSPC histogram with $t_1 = 320$. The circular mean estimate (yellow) and the standard mean estimate (red) superimposed.

Importantly, the sketch formed using the moment in (11) is equivalent to the complex valued ECF sketch $\mathbf{z}_n = \frac{1}{n}\sum_{j=1}^{n} e^{\mathrm{i}\omega x_j}$ sampled at $\omega = \frac{2\pi}{T}$, decoupled into both its real and imaginary components. In fact, the estimate $\hat{t}_1$ in (12) is the optimal estimator to the compressive ECF sketch detailed in (9) (see Appendix). Principally, we only need to store and transfer 2 values to accurately estimate the depth location of the object or surface, without the requirement to recourse to the original photon time-stamped data. For the remainder of this section, we generalize the approach of forming a sketch in (8) of arbitrary size and sampling the ECF at multiple frequencies $[\omega_i]_{i=1}^m$. This will enable us to obtain statistically efficient estimates for the single surface case and to solve more complex lidar scenes including several surfaces with varying intensities where more salient information of the observation model is required.

B. Sampling the ECF

Recall that the observation model π in (2) is discretized over the interval $[0, T-1]$ which we can consider to be a sufficient sampling if the distribution in (2) is approximately

bandlimited. As a result, the characteristic function $\Psi_\pi(\omega)$ has a finite basis characterized by the set of frequencies

$$\left\{ \frac{2\pi j}{T} \;\middle|\; j \in [0, T-1] \right\}. \tag{13}$$

We can generalise the approach from Section III-A by sampling multiple frequencies from the finite basis in order to construct the ECF sketch. As is the case for the circular mean, the frequencies $\omega = \frac{2\pi j}{T}$ for $j \in [0, T-1]$ correspond to the zeros in the sinc function associated with the background pdf π_b seen in (10). We can therefore construct a sketch of arbitrary dimension m that is also *blind* to photons originating from background sources by avoiding the zero frequency $\omega = 0$ of the finite basis. As a result, we define the set of orthogonal frequencies by

$$\Omega \coloneqq \left\{ \omega_j = \frac{2\pi j}{T} \;\middle|\; j \in [0, T-1] \right\}. \tag{14}$$

We coin this set the *orthogonal frequencies* as it defines regions over the interval of the observation model's characteristic function where the signal's contribution is orthogonal to the background's contribution.

1) Sampling Schemes: In order to construct a sketch, we are ultimately interested in retaining sufficient salient information of the characteristic function Ψ_π such that we can identify and estimate the unique location and intensity parameters θ of the observation model $\pi(x \mid \theta)$ defined in (2). It was discussed in Section II that the CF of a probability distribution decays in frequency, i.e. $\Psi_\pi(\omega) \to 0$ as $\omega \to \infty$. Furthermore, as the observation model is discretized over the interval, we assume that the characteristic function of the observation model is approximately band-limited. A natural sampling scheme would therefore be to sample the first m frequencies of the orthogonal frequencies Ω to capture the maximum energy of the CF. In other words, we could truncate the CF of the observation model whilst avoiding the zero frequency.

Alternatively, in [24], [25], provable guarantees for estimating mixture of Gaussian models have been provided, under certain conditions based on random sampling (cf. compressive sensing [34]) of the CF. It is understood that the higher frequencies of the CF may provide further information to help discriminate distributions that are close in probability space. Moreover, if the CF decays slowly in frequency then the energy of the CF will be spread more throughout the set of orthogonal frequencies. We therefore provide an alternative sampling scheme whereby we randomly sample the set of orthogonal frequencies with respect to some sampling law Λ. In a similar design to the frequency sampling pattern proposed in [35], we sample the orthogonal frequencies by

$$(\omega_1, \omega_2, \ldots, \omega_m) \sim \Lambda_{\hat{h}}, \tag{15}$$

where $\Lambda_{\hat{h}} \propto \hat{h}$. To formalize, we consider the follow sampling schemes in order to construct our ECF sketches:

1) Truncated Orthogonal Sampling: Sample the first m frequencies i.e $j = 1, 2, \ldots, m$ from Ω.
2) Random Orthogonal Sampling: Sample the set of frequencies randomly, governed by the distributing law $\Lambda_{\hat{h}}$.

Depending on the circumstances of the lidar device we might expect one or the other sampling scheme to perform better.

C. Statistical Estimation

Once the ECF sketch is constructed using either sampling scheme, we must estimate the parameters θ of the observation model $\pi(x \mid \theta)$ solely from the sketch $\mathbf{z}_n$. In general, there is no closed form expression to estimate θ from the sketch of arbitrary size as is the case for the circular mean estimate in (12).

It is well documented in the ECF and GeMM literature, e.g. [21], [22], [36], that a complex valued ECF sketch $\mathbf{z}_n$ of size m, computed over a finite dataset $\mathcal{X} = \{x_1, \ldots, x_n\}$, satisfies the central limit theorem. Formally, a sketch $\mathbf{z}_n \in \mathbb{C}^m$ converges asymptotically to a Gaussian random variable

$$\mathbf{z}_n \xrightarrow{\text{dist}} \mathcal{N}\big(\Psi_\pi, n^{-1}\Sigma_\theta\big), \tag{16}$$

where $\Sigma_\theta \in \mathbb{C}^{m \times m}$ has entries $(\Sigma_\theta)_{ij} = \Psi_\pi(\omega_i - \omega_j) - \Psi_\pi(\omega_i)\Psi_\pi(-\omega_j)$ for $i, j = 1, 2, \ldots, m$.

The asymptotic normality result in (16) naturally leads to a sketch maximum likelihood estimation (SMLE) algorithm that consists of minimising the following

$$\arg\min_\theta \;\; \frac{m}{2} \log\det(\Sigma_\theta) + n(\mathbf{z}_n - \mathbf{z}_\theta)^T \Sigma_\theta^{-1} (\mathbf{z}_n - \mathbf{z}_\theta), \tag{17}$$

where for convenience we denote $\mathbf{z}_\theta = \big[\Psi_\pi(\omega_j)\big]_{j=1}^m$. For an observation model consisting of K surfaces and a general impulse response function h, recall that

$$\mathbf{z}_\theta = \left[\sum_{k=1}^{K} \alpha_k \hat{h}(\omega_j) e^{\mathrm{i}\omega_j t_k} \right]_{j=1}^{m} \tag{18}$$

and $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_K, t_1, \ldots, t_K)$. Note that we have dropped the sinc function on the assumption that we are using one of the proposed sampling schemes. Minimising (17) is equivalent to minimising the compressive GeMM objective function defined in (9) with the weighting function chosen to be $\mathbf{W} = \Sigma_\theta^{-1}$. The weighting matrix $\mathbf{W} = \Sigma_\theta^{-1}$ is asymptotically optimal in the sense that it minimises the variance of the estimator $\hat{\theta}$ from the sketch sketch $\mathbf{z}_n$ [21].

In practice Σ_θ is θ dependent as it is a function of the underlying parameters θ that are to be estimated. There are various well established methods in the GeMM and ECF literature [20], [21] that tackle the difficulty of approximating Σ_θ and estimating θ simultaneously. In [31], they use the K-L method which iteratively estimates Σ_θ and θ in a two stage procedure by fixing and updating one at a time. Some particular methods [37] fix Σ_θ after only a few iterations of the K-L approach to reduce the computational complexity of the algorithm, although this typically comes at the cost of introducing sample bias [38]. Occasionally, the covariance matrix is set throughout to be the identity, $\Sigma_\theta = I$, reducing (17) to a standard least squares optimization but this generally results in a less statistically efficient estimator $\hat{\theta}$ [20].

In this paper, we use the method of estimating Σ_θ and θ simultaneously at each iteration. This approach is commonly referred to as Continuous Updating Estimator (CUE) [37] and obtains estimates that do not produce sample bias like the two-step K-L approach [38] and can often lead to more statistically efficient estimators [20]. However, the SMLE

method is not restricted to the CUE and in certain situations practitioners may choose to sacrifice unbiased and efficiently optimal estimators for a reduced computational complexity by considering the other methods discussed.

The optimisation problem in (17) is also typically non convex and can suffer from spurious local minima. For the case when there is only a single surface, we initialise the SMLE algorithm using the analytic circular mean solution in (12) with minimal added computational overhead. From our experience with synthetic and real data, the circular mean estimate generally initialises the SMLE algorithm within the basin of the global minima, hence the issues associated with non-convex optimization are circumvented. For the case of multiple surfaces, we form a coarse uniform grid across $[0, T-1]^K$ and initialise at the smallest SMLE loss.

D. Central Limit Theorem

One of the main advantages of the SMLE lidar approach from (16) is that even at low photon levels (i.e. small n), the SMLE estimates quickly follow the central limit theorem (CLT) and provide a good approximation of its expectation. In contrast, the TCSPC histogram used for many estimation methods, discussed in Section I, is a poor approximation to its expectation as each time stamp bin t has only a small number of photons. Thus efficient processing of the full histogram data requires careful consideration of the underlying Poisson statistics [39].

This is illustrated in Figure 3 which shows four separate histograms of the error $(\hat{t} - t_1)$ for increasing photon count n, along with the asymptotic Gaussian distribution from (16). The estimate $\hat{t}$ was obtained from a real valued sketch of size 2 ($m = 1$) using the circular mean estimate in (12). The simulated data was the same as the motivation example in Section III-A where a Gaussian IRF with $\sigma = 15$ was used. The SBR was set at 1 and the total number of time-stamps was $T = 1000$. The total photon count varied from $n = 10$ to $n = 10000$ increasing by a factor of 10 each time. For each photon count, we estimated the location parameter t_1 a total of 1000 times where the data $\mathcal{X} = \{x_i\}_{i=1}^m$ was simulated independently for each trial.

Even at extremely low photon counts of $n = 10$, the error $(\hat{t} - t_1)$ can be reasonably approximated by a Gaussian random variable centred around 0. This suggests that the estimate $\hat{t}$ quickly satisfies the central limit theorem with respect to the photon count n. In the large photon regime ($n = 10000$), the estimation error is concentrated tightly around zero and mostly contained within 5 time stamps. These results suggest that the sketched lidar CLT results of (16) hold even for low photons levels, hence the SMLE loss in (17) is a well-justified loss to minimise.

E. Statistical Efficiency

In this section, we calculate the theoretical statistical efficiency of the sketched lidar estimates, θ, that parametrize the observation model $\pi(x \mid \theta)$ in (2), and compare them with the estimates obtained using the full data (i.e no compression)

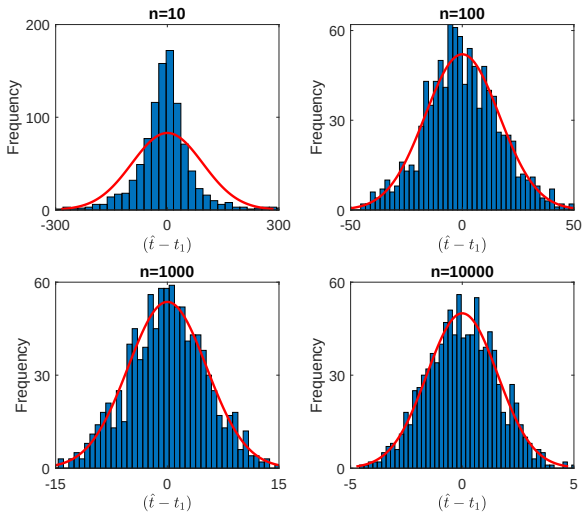

Fig. 3: Histograms of the $(\hat{t} - t_1)$ for increasing photon count n where the sketched lidar estimate (circular mean) is denoted by $\hat{t}$.

using the relative error percentage. The relative error percentage, which will be defined later, is a key metric allowing us to quantify the relative loss of information given a sketch of size m from a statistical point of view.

Statistical efficiency is a measure of the variability or quality of an unbiased estimator $\hat{\theta}$ [40]. The Cramér-Rao bound gives a lower bound on the mean squared error of $\hat{\theta}$ [41] and therefore provides a best case scenario on the variability of the parameter estimates. Given the observation model $\pi(x \mid \theta)$ and the corresponding Fisher information matrix (FIM), defined as

$$\mathcal{I}_{\text{data}}(\theta) \coloneqq n\mathbb{E}\left[\left(\frac{\partial \log \pi(x \mid \theta)}{\partial \theta}\right)^2\right], \tag{19}$$

then the optimal Cramér-Rao mean squared error, in terms of the full data, is defined as

$$\text{RMSE}_n \coloneqq \sqrt{\sum_{k=1}^{2K} [\mathcal{I}_{\text{data}}(\theta)^{-1}]_{\{kk\}}}. \tag{20}$$

Equivalently, we can compute the FIM for the sketched case using the normality result stated in (16), where the FIM of a multivariate Gaussian distribution [41] is defined as

$$\left(\mathcal{I}_{\text{sketch}}(\theta)\right)_{ij} \coloneqq n \frac{\partial \mathbf{z}_\theta}{\partial \theta_i} \Sigma_{\theta_0}^{-1} \frac{\partial \mathbf{z}_\theta}{\partial \theta_j}, \tag{21}$$

where $\mathbf{z}_\theta$ is the sketch defined in (18). Similarly, we define the optimal sketched Cramér-Rao mean squared error as

$$\text{RMSE}_m \coloneqq \sqrt{\sum_{k=1}^{2K} [\mathcal{I}_{\text{sketch}}(\theta)^{-1}]_{\{kk\}}}. \tag{22}$$

To quantify the statistical efficiency of an estimate obtained from a real valued sketch of size $2m$, we use the relative error percentage (REP) metric which compares the optimal sketch

root mean squared error RMSE_m with the corresponding full data root mean squared error RMSE_n, defined by

$$\text{REP} \coloneqq 100\left(\frac{\text{RMSE}_m - \text{RMSE}_n}{\text{RMSE}_n}\right). \tag{23}$$

Notably, the FIM of the sketched statistic in (21) scales with n, hence the REP metric is independent of the photon count.

We compare the statistical efficiency of the sketched lidar estimates to the alternative compression technique of coarse binning [8] discussed in Section I. The coarse binning approach can be seen to be equivalent to constructing a summary statistic

$$\tilde{\mathbf{z}}_n = \sum_{i=1}^{n} \left\{ \mathbb{1}_{[(j-1)\Delta_{\tilde{m}}, j\Delta_{\tilde{m}}]}(x_i) \right\}_{j=1}^{\tilde{m}}, \tag{24}$$

where $\Delta_{\tilde{m}} = \lceil \frac{T}{\tilde{m}} \rceil$ denotes the down-sampling factor, $\tilde{m}$ denotes the number of measurements equivalent to the real-valued sketch size (i.e. $\tilde{m} = 2m$) and $\mathbb{1}_{[t_i, t_i+\Delta_{\tilde{m}}]}(x)$ is the indicator function defined as

$$\mathbb{1}_{[(j-1)\Delta_{\tilde{m}}, j\Delta_{\tilde{m}}]}(x) \coloneqq \begin{cases} 1 & \text{if } \; x \in \big[(j-1)\Delta_{\tilde{m}}, j\Delta_{\tilde{m}}\big), \\ 0 & \text{Otherwise.} \end{cases} \tag{25}$$

Once the coarse binning sketch has been constructed, traditional estimation methods, for e.g log-matched filtering [42] or expectation maximization [43], can be employed on the sketch to estimate the parameters of the observation model.

Lidar scenes typically have only 0,1 or 2 reflectors in the scene, therefore in the following experiments we only consider the case where $K = 1, 2$. Moreover, we choose the setting of the lidar scene (e.g. binning resolution, peak location, intensity) to best replicate a realistic setting as seen in Section IV-C. In each experiment, we consider two different impulse response functions (IRF), exhibiting both a short and heavy-tail. Figure 4 depicts the contrasting IRFs and the magnitude of their corresponding characteristic functions, $\Psi_{\pi_s}(\omega) = \hat{h}(\omega)e^{\mathrm{i}\omega t}$. We evaluate the statistical efficiency of the sketched and coarse binning estimate using the REP as a function of the number of real measurements $2m$ and examine both the random and truncated orthogonal sampling schemes discussed in Section III-B1.

1) One Surface: We first evaluate the REP for a single peak case positioned at $t_1 = 430$, a binning resolution of $T = 1000$. We consider both low and high background photon count levels, where the signal to background ratio (SBR) was set at 10 and 1, respectively. Figure 5 shows the REP metric as a function of the number of real measurements $2m$ for the truncated orthogonal (blue), random orthogonal (red) and coarse binning (orange) compression techniques, where the high (SBR=10) and low (SBR=1) background photon levels are denoted by a solid and dashed line, respectively. The top and bottom plots depict the short and heavy-tailed IRF, accordingly.

We first observe that both sketched lidar sampling schemes approach 0% as the real measurements increases and only a modest number of measurements are needed to obtain a low REP. In contrast, the coarse binning approach exhibits a slow convergence REP and remains high throughout the measurement range. Importantly, we see that the different sketch sampling schemes outperform each other depending on the tail of the IRF and hence the rate of decay of the CF. For instance, the truncated scheme produces a lower REP for the short-tailed IRF, while the random sampling scheme achieves a quicker convergence and a significantly lower REP throughout the measurement range for the heavy-tailed IRF. This can be explained by Figure 4, the CF of the short-tailed IRF has the majority of its energy contained within the first few ($m = 10$) frequencies, while the CF of the heavy-tailed IRF has its energy spread more throughout its frequency.

2) Two Surfaces: We now evaluate the REP for a two peak case positioned at $(t_1, t_2) = (320, 570)$, a time stamp resolution of $T = 1000$. The intensity of the two peaks is given by 75% and 25%, respectively, simulating an object that is positioned behind a semi-transparent surface. We simulate both low and high background levels, where the SBR was

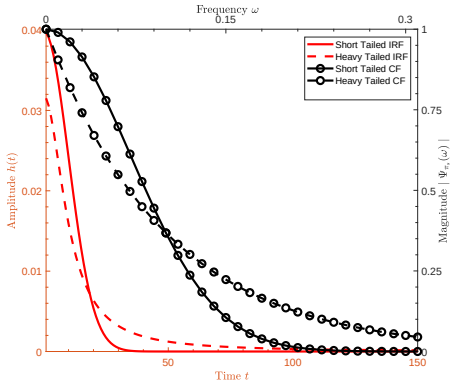

Fig. 4: A figure showing the CF (black) of a short (solid) and heavy (dashed) tailed impulse response function (red).

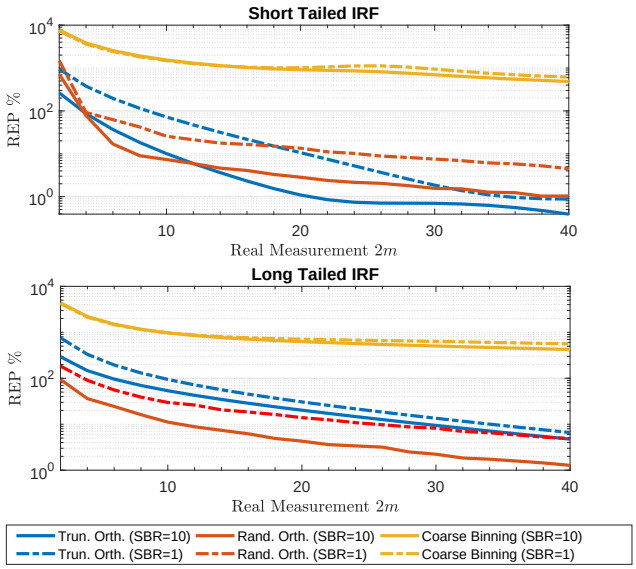

Fig. 5: A figure showing the REP as a function of the number of real measurements ($2m$) for a single peak lidar scene. Top: short-tailed impulse response function. Bottom: heavy-tailed impulse response function.

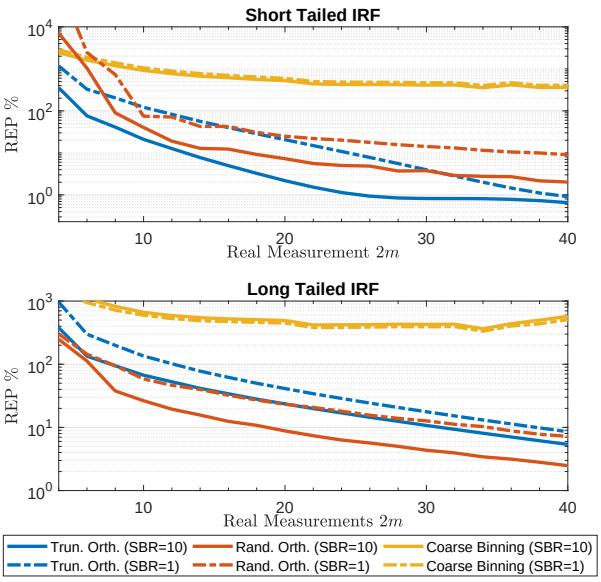

Fig. 6: A figure showing the REP as a function of the number of real measurements ($2m$) for a lidar scene with 2 surfaces. Top: short-tailed impulse response function. Bottom: heavy-tailed impulse response function.

again set at 10 and 1, respectively. Figure 6 shows the REP metric as a function of the number of real measurements $2m$ for the truncated orthogonal (blue), random orthogonal (red) and coarse binning (orange) compression techniques, where the high (SBR=10) and low (SBR=1) background photon levels are denoted by a solid and dashed line, respectively. The top and bottom plots depict the short and heavy-tailed IRF, accordingly. We see the same pattern as the single surface case where the REP remains high for the coarse binning compression technique while, in contrast, the sketched lidar converges towards a relatively low REP in a modest number of measurements. We again observe that the truncated scheme performs best on a fast decaying CF, while the random sampling scheme outperforms the truncated counterpart when there is a slow decaying CF. The doubling of the dimension of the parameter θ by estimating two peaks and intensities, does not have a significant impact on the required number of measurements needed to achieve a relatively low REP. For instance in the high SBR (solid) scenario, the truncated orthogonal sampling scheme requires 20 real measurements($m = 10$) to achieve a REP less than 1% for the unimodal case compared with a requirement of 24 real measurements ($m = 12$) to achieve the same level of REP for the bimodal case.

These theoretical results on the statistical efficiency of the lidar sketch show that only a moderate sketch size is needed to achieve negligible loss of information. The results are based on the asymptotic normality property discussed in (16), and we will see in Section IV-B that in practice this normality result holds even for small photon counts of $n = 10$.

IV. Experiments

A. Experimental set up

In this section, we evaluate our compressive lidar framework on synthetic and real data with increasingly complex scenes. Our method is compared with classical algorithms working on the full data space (i.e no compression) namely log-matched filtering [42] and expectation maximization (EM) [43]. Moreover, we also compare our results to the alternative compression technique of coarse binning [8] discussed in Section I and (24). Both the log-matched filtering and EM algorithms estimate the location parameters using the full data and therefore the results obtained from these methods set a benchmark to the estimation accuracy when no compression takes place. For sake of fair comparison, we use the real valued sketch in all the subsequent results, such that the number of real measurements is equivalent to $2m$.

1) Processing: Restoration of depth imaging of single-photon lidar consists of estimating a 3D point cloud from a lidar data cube containing the number of photons $n_{i,j,t}$ in pixel (i, j) at time stamp t, where $i \in [1, \dots, N_r], j \in [1, \dots, N_c]$ and $t \in [0, T-1]$. We denote the average photon count for each pixel by $\bar{n}$ and process each pixel (i, j) of the data cube and estimate the true location and intensity parameter, denoted t_1 and α, respectively. A data driven impulse response is given for each dataset and we can obtain the characteristic function of the IRF by using (10).

2) Evaluation Metrics: Two different error metrics are used to evaluate the performance of our proposed sketched lidar framework. We consider the root mean squared error (RMSE) between the reconstructed image and the ground truth. Given that $t_{i,j,k}$ is the location of the kth peak in pixel (i, j) and $\hat{t}_{i,j,k}$ the estimated counterpart, then the root mean squared error of the reconstructed image is

$$\text{RMSE} := \sqrt{\frac{1}{KN_rN_c}\sum_{i=1}^{N_r}\sum_{j=1}^{N_c}\sum_{k=1}^{K}\left(t_{i,j,k} - \hat{t}_{i,j,k}\right)^2}. \quad (26)$$

Secondly, we compare the percentage of true detections $F_{\text{true}}(\tau)$ with a resolution τ. That is, a true detection occurs if $|t_{i,j,k} - \hat{t}_{i,j,k}| \leq \tau$ [5]. When plotted as a function of τ, this metric quantifies the variability of the estimate.

The compression of both the sketched lidar and coarse binning approach is measured in terms of the dimension reduction achieved by the statistic with respect to the raw TCSPC data and is quantified by the metric $\max\{\frac{2m}{T}, \frac{2m}{n}\}$, which is dependent on the dimensions, T and n, of the lidar scene and where the number of real measurements ($2m$) is used for sake of fair comparison.

B. Synthetic Data

We evaluate the sketched lidar framework on a synthetic dataset simulating a pixel in a scene which consists of a single peak response. We chose the parameters that best replicated a realistic lidar scene and that were akin to the real datasets which will be discussed in IV-C. Therefore, we set the binning resolution at $T = 500$, and impulse response was generated with a true Gaussian function where $\sigma = 10$, resulting in 2%

standard deviation with respect to the time stamp resolution. The SBR was set at 1 which reflects a moderately noisy and challenging scene, and results in $\alpha_0 = \alpha_1 = 0.5$. We ran a Monte-Carlo simulation with 1000 trials to evaluate and compare the performance of our sketched lidar framework for different photon counts n and varying sketch sizes m.

For each trial, we uniformly chose $t_1 \sim \mathcal{U}(0, 499)$, and estimated $\hat{t}_{\text{sketch}}$ for the sketched lidar approach and $\hat{t}_{\text{coarse}}$ for the alternative compression technique of coarse binning. For point of reference, we estimated the log-matched filter estimate, $\hat{t}_{\text{LMF}}$, which represents the estimate over the full data (i.e. no compression). We varied the total number of real measurements between 2 ($m = 1$) and 20 ($m = 10$) for both the sketched lidar and coarse binning method and increased the photon count from $n = 10$ to $n = 10000$ by a factor of 10 each time. The solid lines in Figure 4 show the IRF h and the magnitude of its corresponding CF, respectively. Here we only show the results for the truncated orthogonal sampling scheme but we observed in practice that the alternative random orthogonal sampling scheme produces similar results.

Figure 7 shows the RMSE as a function of the number of real measurements $2m$ for the four increasing photon counts n. For both compression techniques, the RMSE decreases on average as m increases. The sketched lidar method (blue) exhibits a significantly smaller RMSE than the alternative course binning technique for all photon counts. Notably, the sketched lidar method converges to a low RMSE even at the low photon regime of $n = 10$, suggesting that the normality result in (16) holds for low photon counts. Our sketched approach even achieves a smaller RMSE than the LMF[1] for $n = 10, 100$ and 1000 which suggests the loss of statistical efficiency is negligible even for small sketch sizes as shown in Section III-E.

Figure 8 shows the percentage of peaks detected as a function of the resolution τ for the different real valued sketch sizes of $2, 10$ and 20. The sketched lidar method performs comparably with the LMF even at low photon counts, achieving approximately 98% of peak detections within only $\tau = 10$ resolutions for a real valued sketch size of 10 ($m = 5$) and a photon count of $n = 10$. Impressively, for the large photon regime of $n = 10000$ the sketched lidar approach estimates 100% of peaks within $\tau = 1$ resolution for all the shown real valued sketch sizes. In comparison the alternative coarse binning method only estimates 8% of peaks within the same resolution $\tau = 1$ for the largest sketch size of $\tilde{m} = 20$.

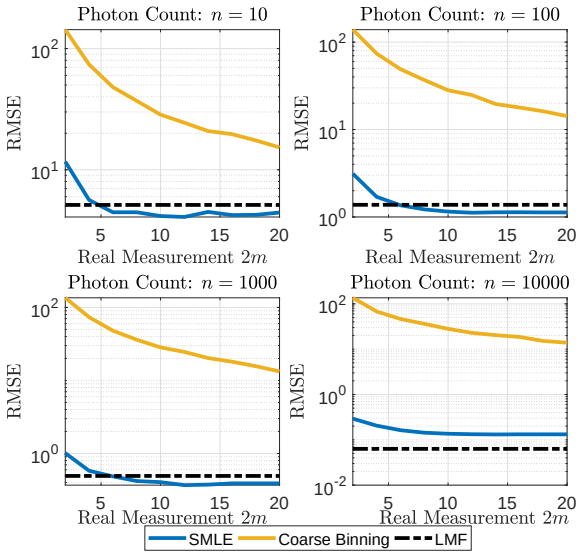

Fig. 7: A figure showing the RMSE as a function of the number of real measurements ($2m$) for both the sketched lidar (blue) and coarse binning (orange) method. The log matched filter (dashed black) is shown for comparison.

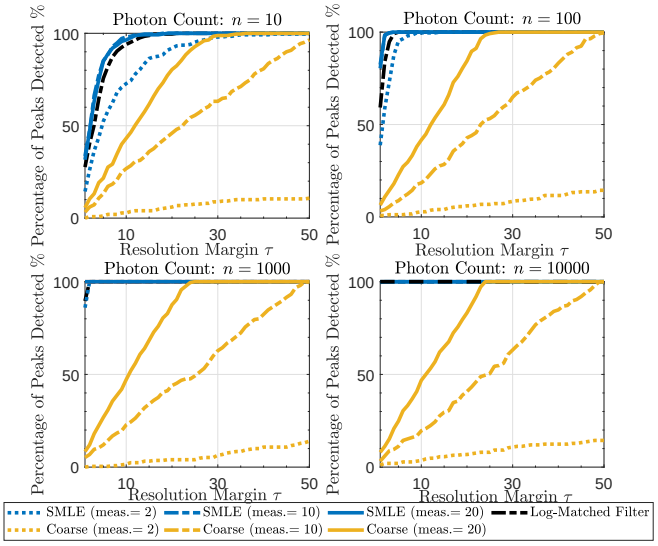

Fig. 8: A figure showing the percentage of peak detected as a function of the time stamp resolution τ for both the sketched lidar and coarse binning method for varying real measurement sizes ($2m$).

C. Real Data

In this section we evaluate our sketched lidar framework on two real datasets of increasing complexity. Namely, a polystyrene head imaged at Heriot-Watt University [5], [28] which consists mostly of a single peak, and a scene where two humans are standing behind a camouflage net, depicted in [7], [44], which contains of 2 objects per pixel with varying intensity.

1) Polystyrene Head: The first scene consists of a polystyrene head placed 40 meters away from the lidar device. The data cube has width and height of 141 pixels, $N_r = N_c = 141$ and a total of $T = 4613$ time stamps. A total acquisition time of 100 milliseconds was used for each pixel resulting in an average photon count of $\bar{n} = 337$ with an SBR of approximately 6.82. The vast majority of pixels consist of a single peak, although there are a minority of pixels around the borders of the head that consist of two peaks. The parameter set to be estimated for each pixel is $\theta = (t, \alpha)$ of dimension 2. We compare our results with the

[1]Log-matched filtering works under the assumption of negligible background noise while here the SBR is 1.

ground truth obtained from the experiment as well as the full data algorithm of log-matched filtering and the coarse binning compression technique. As the log-matched filter is the maximum likelihood estimation of a single peak, we assume each pixel has one surface for the sake of comparison. The coarse binning approach is computed using log-matched filtering once the data cube is down-sampled.

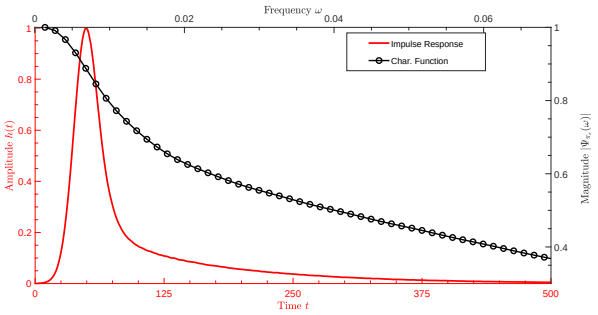

Fig. 9: A figure showing the CF (black) of the data driven impulse response function (red) of the polystyrene head dataset.

The data driven impulse response function and its corresponding CF obtained from (10), are shown in Figure 9. We only present the results for the truncated orthogonal sampling scheme, from Section III-B1, but we observed in practice that the alternative random orthogonal sampling scheme produces similar results. We initialise the sketched lidar algorithm using the analytic circular mean solution in (12).

Figure 10 shows the reconstructed images of the sketched lidar, coarse binning and log matched filter approaches, as well as the ground truth image. We first notice that our sketched lidar method sufficiently reconstructs the polystyrene head scene for all sketch sizes, even for the circular mean estimate ($m = 1$) in (a). In contrast, the coarse binning approach fails for all corresponding measurements $\tilde{m}$ with significant staircase artifacts arising. Figure 11 shows the RMSE, in comparison to the ground truth, as a function of the number of real measurements ($2m$). We observe that our sketched lidar method produces a smaller RMSE as the measurement size increases and achieves a smaller RMSE than the LMF approach for larger measurements. In comparison, the coarse binning method obtain estimates that produce a large RMSE consistently throughout. As such, this suggests that our sketched lidar approach does not compromise reduced resolution in favour of compression which is very apparent in the coarse binning method.

2) Humans Behind Camouflage : The second scene consists of two humans standing behind a camouflage net approximately 320 metres away from the lidar device. Further details can be found of the scene in [44], [45]. The data cube has width and height of 32 pixels, $N_r = N_c = 32$ and a total of $T = 153$ time stamps. A total acquisition time of 5.6 milliseconds was used for each pixel resulting in an average photon count of $\bar{n} = 871$ with an approximate SBR of 2.35. The vast majority of pixels have 2 surfaces (the camouflage net and a human) where the net (first peak) accounts for the biggest intensity. The parameter set to be estimated for each pixel is $\theta = (t_1, t_2, \alpha_1, \alpha_2)$ of dimension 4. We compare our

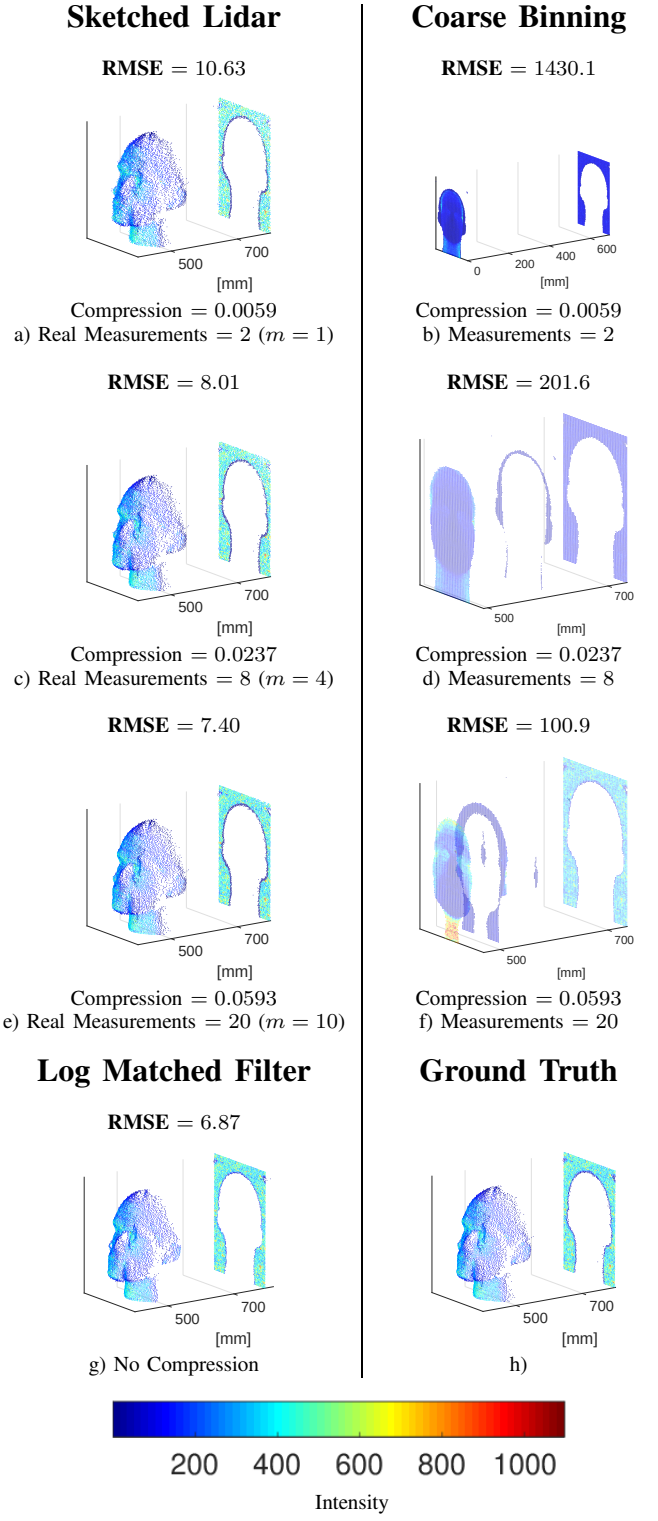

Fig. 10: A figure showing the face dataset lidar reconstructions of the sketched lidar and coarse binning method for the real valued measurement size $2, 8, 20$. Both the log matched filter reconstruction and the ground truth image are given for comparison.

results with the full data EM algorithm as well as the coarse binning compression technique. For this experiment, the coarse binning algorithm uses the EM estimate once the data cube has

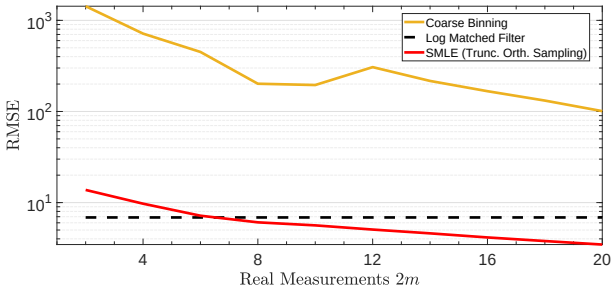

Fig. 11: A figure showing the RMSE as a function of the number of real measurements ($2m$) for the polystyrene head dataset. The random orthogonal sketched lidar (red) and coarse binning method (orange) are compared with the full data estimation via log matched filtering (dashed black).

been down-sampled as the log-match filtering algorithm is only applicable to single peak cases. Due to the lack of a ground truth, we compare the reconstructions of the camouflage scene to the full data EM algorithm reconstruction and equate the relevant compression of both the sketched lidar framework and the coarse binning technique. The data driven impulse response function h and its corresponding CF obtained from (10), are shown in Figure 12. Again, we only present the results for the truncated orthogonal sampling scheme, from Section III-B1, but we observed in practice that the alternative random orthogonal sampling scheme produces similar results. We uniformly sampled 10 starting points for each of peak t_1 and t_2 and initialised with the smallest sketched cost function from (9).

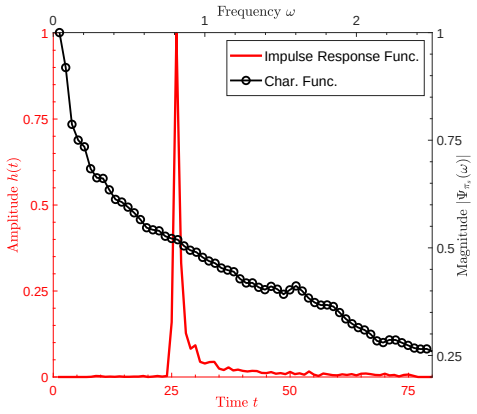

Fig. 12: A figure showing the CF (black) of the data driven impulse response function (red) of the camouflage dataset.

Figure 13 shows the reconstructed images of the sketched lidar, coarse binning and EM algorithm methods. Evidently, the reconstruction of our sketched lidar approach becomes better as the number of real measurements ($2m$) increases, for instance the torso of the human positioned near 600 cm has greater clarity in sketch size 20 compared to sketch size 4 where more spurious peaks are detected. However, the sketched lidar reconstruction for $m = 2$ is still sufficient in comparison to the EM reconstruction in (g), while in contrast the coarse binning method fails to reconstruct either human for the corresponding number of measurements. The coarse binning method once again suffers from the stair case effect as seen by the lack of width of the first human standing at position 200 cm in (f). Furthermore, the compression due to the coarse binning results in poor depth accuracy as seen by the position of the camouflage net in reconstruction (b) which has a disparity of approximately 120 cm in comparison to the EM reconstruction. Once again, this suggests that our sketched lidar approach does not compromise reduced resolution in favour of compression which is very apparent in the coarse binning method.

V. Conclusion

In this paper, we proposed a novel sketching solution to handle the major data processing bottleneck of single-photon lidar caused by the fine resolution of modern high rate, high resolution ToF image sensors. Our approach involved sampling the characteristic function of the observation model to form online statistics that have dimensionality proportional to the number of parameters of the model. Furthermore, we developed an efficient sketching algorithm, inspired by ECF estimation techniques, which has space and time complexity that fundamentally scales with the size of the sketch m, and is independent of both photon count and depth resolution. Two sampling schemes are proposed that sample in regions of the characteristic function that are *blind* to photons originating from background sources. As a result, our method obtains estimates of the location and intensity parameters that are unbiased.

Our novel sketch based acquisition removes the trade-off between depth resolution and data transfer complexity that is apparent in existing methods. Here we have only considered a simple pixel-wise depth estimate method in the form of the sketched MLE. It should be straightforward to incorporate the sketched statistics into more sophisticated state-of-the-art reconstruction algorithms, such as the real-time 3D algorithm in [7]. However, we leave this for future work. Another line of future work would be to use the sketch statistics for other algorithmic purposes such as target detection and multi-reflection detection.

Acknowledgements

This work was supported by the ERC Advanced grant, project C-SENSE, (ERC-ADG-2015-694888). Mike. E. Davies is also supported by a Royal Society Wolfson Research Merit Award. The authors would like to thank the single-photon group at HWU (https://single-photon.com) for the use of the datasets used in Section IV-C. The polystyrene head dataset in IV-C1 and the camouflage dataset in IV-C2 were obtained from [5] and [7], respectively.

References

[1] J. Hecht, "Lidar for self-driving cars," *Opt. Photon. News*, vol. 29, no. 1, pp. 26–33, Jan 2018.

[2] J. Rapp, J. Tachella, Y. Altmann, S. McLaughlin, and V. K. Goyal, "Advances in single-photon lidar for autonomous vehicles: Working principles, challenges, and recent advances," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 62–71, 2020.

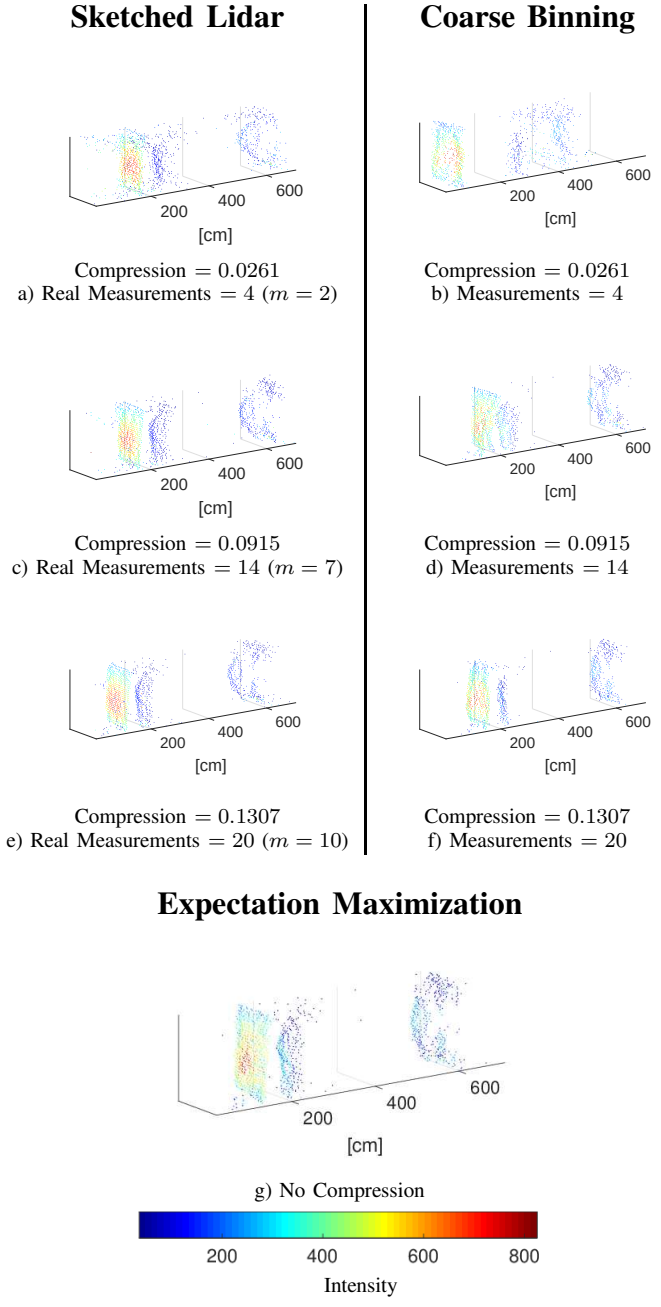

Fig. 13: A figure showing the camouflage dataset lidar reconstructions of the sketched lidar and coarse binning method for the real valued measurement size ($2m$) of $2, 8, 20$. Both the log matched filter reconstruction and the ground truth image are given for comparison.

[3] J. Gao, J. Sun, J. Wei, and Q. Wang, "Research of underwater target detection using a slit streak tube imaging lidar," in *2011 Academic International Symposium on Optoelectronics and Microelectronics Technology*, 2011, pp. 240–243.

[4] M. Pierzchała, P. Giguère, and R. Astrup, "Mapping forests using an unmanned ground vehicle with 3D lidar and graph-slam," *Computers and Electronics in Agriculture*, vol. 145, pp. 217 – 225, 2018.

[5] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, S. McLaughlin, and J. Tourneret, "Bayesian 3D reconstruction of complex scenes from single-photon lidar data," *SIAM Journal on Imaging Sciences*, vol. 12, pp. 521–550, 03 2019.

[6] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, "Single-photon three-dimensional imaging at up to 10 kilometers range." *Optics express*, vol. 25 10, pp. 11 919–11 931, 2017.

[7] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J. Tourneret, and S. McLaughlin, "Real-time 3D reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," *Nature communications*, vol. 10, no. 1, pp. 1–6, 2019.

[8] R. K. Henderson, N. Johnston, H. Chen, D. D. Li, G. Hungerford, R. Hirsch, D. McLoskey, P. Yip, and D. J. S. Birch, "A 192×128 time correlated single photon counting imager in 40nm CMOS technology," in *ESSCIRC 2018 - IEEE 44th European Solid State Circuits Conference (ESSCIRC)*, 2018, pp. 54–57.

[9] X. Ren, P. W. R. Connolly, A. Halimi, Y. Altmann, S. McLaughlin, I. Gyongy, R. K. Henderson, and G. S. Buller, "High-resolution depth profiling using a range-gated CMOS SPAD quanta image sensor," *Opt. Express*, vol. 26, no. 5, pp. 5541–5557, Mar 2018.

[10] R. J. Walker, J. A. Richardson, and R. K. Henderson, "A 128×96 pixel event-driven phase-domain $\delta\sigma$-based fully digital 3D camera in 0.13μm CMOS imaging technology," in *2011 IEEE International Solid-State Circuits Conference*, 2011, pp. 410–412.

[11] F. M. Della Rocca, H. Mai, S. W. Hutchings, T. Al Abbas, A. Tsiamis, P. Lomax, I. Gyongy, N. A. W. Dutton, and R. K. Henderson, "A 128 × 128 SPAD dynamic vision-triggered time of flight imager," in *ESSCIRC 2019 - IEEE 45th European Solid State Circuits Conference (ESSCIRC)*, 2019, pp. 93–96.

[12] F. Mattioli Della Rocca, H. Mai, S. W. Hutchings, T. A. Abbas, K. Buckbee, A. Tsiamis, P. Lomax, I. Gyongy, N. A. W. Dutton, and R. K. Henderson, "A 128 × 128 SPAD motion-triggered time-of-flight image sensor with in-pixel histogram and column-parallel vision processor," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 7, pp. 1762–1775, 2020.

[13] S. W. Hutchings, N. Johnston, I. Gyongy, T. Al Abbas, N. A. W. Dutton, M. Tyler, S. Chan, J. Leach, and R. K. Henderson, "A reconfigurable 3-D-stacked SPAD imager with in-pixel histogramming for flash lidar or high-speed time-of-flight imaging," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 11, pp. 2947–2956, 2019.

[14] C. Zhang, S. Lindner, I. M. Antolović, J. Mata Pavia, M. Wolf, and E. Charbon, "A 30-frames/s, 252 × 144 SPAD flash lidar with 1728 dual-clock 48.8-ps tdcs, and pixel-wise integrated histogramming," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1137–1151, 2019.

[15] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[16] A. Kadambi and P. T. Boufounos, "Coded aperture compressive 3-D lidar," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 1166–1170.

[17] R. Tobin, Y. Altmann, X. Ren, A. McCarthy, R. Lamb, S. McLaughlin, and G. S. Buller, "Comparative study of sampling strategies for sparse photon multispectral lidar imaging: Towards mosaic filter arrays," *Journal of Optics*, vol. 19, p. 094006, 09 2017.

[18] A. Halimi, P. Ciuciu, A. Mccarthy, S. Mclaughlin, and G. S. Buller, "Fast adaptive scene sampling for single-photon 3D lidar images," in *IEEE CAMSAP 2019 - International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Le Gosier (Guadeloupe), France, Dec. 2019.

[19] I. Maksymova, C. Steger, and N. Druml, "Review of lidar sensor data acquisition and compression for automotive applications," in *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 2, no. 13, 2018, p. 852.

[20] L. P. Hansen, "Large sample properties of generalized method of moments estimators," *Econometrica*, vol. 50, no. 4, pp. 1029–1054, 1982.

[21] A. Hall, *Generalized Method of Moments*, 11 2007, pp. 230 – 255.

[22] A. Feuerverger and A. Mureika, "The empirical characteristic function and its applications," *The Annals of Statistics*, vol. 5, no. 1, pp. 88–97, 1977.

[23] A. Feuerverger and P. McDunnough, "On the efficiency of empirical characteristic function procedures," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 43, no. 1, pp. 20–27, 1981.

[24] R. Gribonval, G. Blanchard, N. Keriven, and Y. Traonmilin, "Statistical learning guarantees for compressive clustering and compressive mixture modeling," *arXiv preprint arXiv:2004.08085*, 2020.

[25] N. Keriven, A. Bourrier, R. Gribonval, and P. Pérez, "Sketching for large-scale learning of mixture models," *Information and Inference: A Journal of the IMA*, vol. 7, no. 3, pp. 447–508, 2018.

[26] M. P. Sheehan, M. S. Kotzagiannidis, and M. E. Davies, "Compressive independent component analysis," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[27] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson, "Bayesian analysis of lidar signals with multiple returns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2170–2180, 2007.

[28] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar waveform-based analysis of depth images constructed using sparse single-photon data," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 1935–1946, 2016.

[29] Y. Altmann and S. McLaughlin, "Range estimation from single-photon lidar data using a stochastic em approach," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1112–1116.

[30] M. Carrasco and J. P. Florens, "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, pp. 797–834, 2000.

[31] A. Feuerverger and P. McDunnough, "On some Fourier methods for inference," *Journal of the American Statistical Association*, vol. 76, no. 374, pp. 379–387, 1981.

[32] E. Lukacs, O. Szász *et al.*, "On analytic characteristic functions," *Pacific J. Math*, vol. 2, no. 4, pp. 615–625, 1952.

[33] S. R. Jammalamadaka and A. Sengupta, *Topics in circular statistics*. world scientific, 2001, vol. 5.

[34] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge university press, 2012.

[35] N. Keriven, A. Bourrier, R. Gribonval, and P. Pérez, "Sketching for large-scale learning of mixture models," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 6190–6194.

[36] L. Hansen, "Large sample properties of generalized method of moments estimators," *Econometrica*, vol. 50, no. 4, pp. 1029–1054, 1982.

[37] L. P. Hansen, J. Heaton, and A. Yaron, "Finite-sample properties of some alternative GMM estimators," *Journal of Business & Economic Statistics*, vol. 14, no. 3, pp. 262–280, 1996.

[38] J. Hausman, R. Lewis, K. Menzel, and W. Newey, "Properties of the cue estimator and a modification with moments," *Journal of Econometrics*, vol. 165, no. 1, pp. 45 – 57, 2011, moment Restriction-Based Econometric Methods.

[39] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-efficient computational 3-D and reflectivity imaging with single-photon detectors," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 112–125, 2015.

[40] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 222, pp. 309–368, 1922.

[41] S.-i. A and H. Nagaoka, *Methods of information geometry*. American Mathematical Soc., 2007, vol. 191.

[42] G. Turin, "An introduction to matched filters," *IRE Transactions on Information Theory*, vol. 6, no. 3, pp. 311–329, 1960.

[43] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[44] A. Halimi, R. Tobin, A. McCarthy, S. McLaughlin, and G. S. Buller, "Restoration of multilayered single-photon 3D lidar images," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 708–712.

[45] R. Tobin, A. Halimi, A. McCarthy, X. Ren, K. J. McEwan, S. McLaughlin, and G. S. Buller, "Long-range depth profiling of camouflaged targets using single-photon detection," *Optical Engineering*, vol. 57, no. 3, p. 031303, 2017.

APPENDIX

Deriving the Circular Mean Estimate From the ECF Estimation:

Given that we have a single frequency $\omega \in \mathbb{R}$, we can define the sketch as $z_n = \frac{1}{n}\sum_{j=1}^{n} e^{\mathrm{i}\omega x_j}$ and the goal is to solve:

$$\hat{\theta} = \arg\min_{\theta} (z_n - \Psi_\pi(\omega))^2. \tag{27}$$

Clearly, (27) is minimised when $\Psi_\pi(\omega) = z_n$ and equating the real and complex components we get:

$$\alpha e^{\frac{(\omega t)^2}{2}} \cos(\omega t) - (1-\alpha)\,\mathrm{sinc}(\frac{\omega T}{2}) = \frac{1}{n}\sum_{j=1}^{n} \cos(\omega x_j) \tag{28}$$

$$\alpha e^{\frac{(\omega t)^2}{2}} \cos(\omega t) = \frac{1}{n}\sum_{j=1}^{n} \sin(\omega x_j). \tag{29}$$

Notably, we can optimally choose the frequency to be $\omega = \frac{2\pi}{T}$ resulting in $\mathrm{sinc}(\frac{\omega T}{2}) = 0$ and thereby ensure the characteristic function is sampled in a region where the background noise is not present. Consequently, dividing (28) by (29) we get

$$\frac{\alpha e^{\frac{(\frac{2\pi t}{T})^2}{2}} \cos(\frac{2\pi t}{T})}{\alpha e^{\frac{(\frac{2\pi t}{T})^2}{2}} \sin(\frac{2\pi t}{T})} = \frac{\sum_{j=1}^{n} \cos(\omega x_j)}{\sum_{j=1}^{n} \sin(\omega x_j)}, \tag{30}$$

resulting in an optimal estimate of

$$\theta^* = \frac{T}{2\pi} \arctan 2\left(\frac{\sum_{j=1}^{n} \cos(\omega x_j)}{\sum_{j=1}^{n} \sin(\omega x_j)}\right). \tag{31}$$